IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Subutai Ahmad et al.

Assignee: Interval Research Corporation

Title: Browser for Use In Navigating A Body Of Information, With Particular Application To Browsing Information Represented by Audiovisual Data

Serial No.: Unknown Filed: Herewith

Examiner: John Miller Group Art Unit: 2611

Attorney Docket No.: IR-002-C1

Milpitas, California
May 29, 2001

Box Patent Application
Assistant Commissioner for Patents
Washington, D.C. 20231

PRELIMINARY AMENDMENT

Sir:

Please enter the following preliminary amendment in the above-identified application. Appended to this Amendment is a document entitled Version with Markings to Show Changes Made, showing the changes made to the specification and claims of this application in this Amendment.

IN THE SPECIFICATION

Please replace the paragraph beginning at page 30, line 24, with the following rewritten paragraph:

The speed control 217 can be used to increase or decrease the apparent display rate with which the primary information is displayed. The speed control

display 217 shows a number that represents the amount by which a normal display rate is multiplied to produce the current apparent display rate, and includes a graphical slider bar that can be used to adjust the apparent display rate. The manner in which the apparent display rate can be changed is described in more detail below.

Please replace the paragraph beginning at page 55, line 7, with the following rewritten paragraph:

FIG. 4 is a flow chart of a method 400, in accordance with this aspect of the invention, for determining whether a first set of information represented by a first set of data of a first type (e.g., audiovisual data) is relevant to a second set of information represented by a second set of data of a second type (e.g., text data). In step 401, a set of data of the second type is derived from the first set of data of the first type. In a typical application of the method 400, step 401 causes a set of text data to be produced from a set of audiovisual data. The set of text data can be produced in any appropriate manner. For example, "production" of the set of text data may be as simple as extracting a pre-existing text transcript (e.g., a closed caption transcript) from the set of audiovisual data. Or, the set of text data can be produced from the set of audio data using a conventional speech recognition method. In step 402,

the derived set of data (of the second type) is compared to the second set of data of the second type to determine the degree of similarity between the derived set of data and the second set of data. One way of making this determination is described in more detail below. In step 403, a determination is made as to whether the first set of data is relevant to the second set of data, based on the comparison of step 402. Typically, a threshold level of similarity (the expression of which depends upon the method used to determine similarity) is specified so that only sets of information that are sufficiently related to each other are identified as related. (This means, when the method 400 is used to generate the related secondary information region 204, that less than the allotted number of secondary information segments - or even no secondary information segments - may be displayed.)

Please replace the paragraph beginning at page 56, line 5, with the following rewritten paragraph:

The degree of similarity can be determined using any appropriate method, such as, for example, relevance feedback. In relevance feedback, a text representation of each segment to be compared (e.g., each audiovisual news story or text story) is represented as a vector, each component of the vector corresponding to a word, the value of each component being the number of occurrences of the word in the segment. (Two words are

considered identical - i.e., are amalgamated for purposes of ascribing a magnitude to each component of the vector representing the textual content of a segment - if the words have the same stem; for example, "play", "played" and "player" are all considered to be the same word for purposes of forming the segment vector.) For each pair of segments, the normalized dot product of the vectors corresponding to the segments is calculated, yielding a number between 0 and 1. The degree of similarity between two segments is represented by the magnitude of the normalized dot product, 1 representing two segments with identical words and 0 representing two segments having no matching words. The use of relevance feedback to determine the similarity between two text segments is well-known, and is described in more detail in, for example, the textbook entitled Introduction to Modern Information Retrieval, by Gerard Salton, McGraw-Hill, New York, 1983, the pertinent disclosure of which is incorporated by reference herein. Relevance feedback is also described in detail in "Improving Retrieval Performance by Relevance Feedback," Salton, G., Journal of the American Society for Information Science, vol. 41, no. 4, pp. 288-297, June 1990 as well as "The Effect of Adding Relevance Information in a Relevance Feedback Environment," Buckley, C. et. al., Proceedings of 17th International Conference on Research and

Development in Information Retrieval, DIGIR 94, Springer-Verlag (Germany), 1994, pp. 292-300, the disclosures of which are incorporated by reference herein.

Please replace the paragraph beginning at page 67, line 17, with the following rewritten paragraph:

The unsummarized text data is aligned with the unsummarized audio data. If the text data has been obtained from the audio data using a speech recognition method, then the alignment of the unsummarized text data with the unsummarized audio data typically exists as a byproduct of the speech recognition method. Otherwise, alignment is accomplished in three steps. First, the unsummarized text data is evaluated to generate a corresponding linguistic transcription network (e.g., a network describing the set of possible phonetic transcriptions). Second, a feature analysis is performed on the audio samples comprising the unsummarized audio data set to create a set of audio feature data. Third, the linguistic transcription network is compared to the set of audio feature data (using Hidden Markov Models to describe the linguistic units of the linguistic transcription network in terms of audio features) to determine the linguistic transcription (from all of the possible linguistic transcriptions allowed by the linguistic transcription network) which best fits the set of audio feature data.

As a result of this comparison, the audio features of the best fit linguistic transcription are correlated with audio features in the set of audio feature data. The audio features of the best fit linguistic transcription can also be correlated with the linguistic units of the linguistic transcription network. The linguistic units of the linguistic transcription network can, in turn, be correlated with the unsummarized text data. As a consequence of these correlations, an alignment of the unsummarized text data with the unsummarized audio data can be obtained. Using the previously determined text summary and the alignment between the text data and audio data, an audio summary can be produced.

IN THE CLAIMS

Please cancel Claims 1-17 and 35-62.

Please amend the claims as follows:

28. (Amended) A system as in Claim 18, wherein the graphical user interface includes a map region for providing a chronological description of the subject matter content of the audiovisual information and for enabling specification of control instructions that enable navigation within the audiovisual information.

Please enter the following new claims:

63. (New) A system as in Claim 18, wherein the audiovisual information is represented at least partially by digital data, the means for displaying further comprising means for displaying digital data.

64. (New) A system as in Claim 18, wherein the audiovisual information is represented at least partially by analog data, the means for displaying further comprising means for displaying analog data.

REMARKS

Claims 1-62 were pending. Claims 1-17 and 35-62 have been canceled. Claim 28 has been amended. Claims 63 and 64 have been added. Allowance of Claims 18-34, 63 and 64 is requested. If the Examiner wishes to discuss any aspect of this application, the Examiner is invited to telephone Applicants' undersigned attorney at (408) 945-9912.

I hereby certify that this correspondence is being deposited with the United States Postal Service as Express Mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, D.C., 20231, on May 29, 2001.
Express Mail Receipt No. EL637958233US

David R. Graham 5-29-01
David R. Graham Date

Respectfully submitted,

David R. Graham
David R. Graham
Reg. No. 36,150
Attorney for Applicants

Version with Markings to Show Changes Made

(Additions are underlined, deletions are enclosed in brackets)

In the specification:

The paragraph beginning at page 30, line 24 has been amended as follows:

The speed control 217 can be used to increase or decrease the apparent display rate with which the primary information is displayed. The speed control display 217 shows a number that represents the amount by which a normal display rate is multiplied to produce the current apparent display [display] rate, and includes a graphical slider bar that can be used to adjust the apparent display rate. The manner in which the apparent display rate can be changed is described in more detail below.

The paragraph beginning at page 55, line 7 has been amended as follows:

FIG. 4 is a flow chart of a method 400, in accordance with this aspect of the invention, for determining whether a first set of information represented by a first set of data of a first type (e.g., audiovisual data) is relevant to a second set of information represented by a second set of data of a second type (e.g., text data). In step 401, a set of data of the second type is derived from the first set of data of the first type. In a typical application of the method 400, step 401 causes a set of text data to be produced from a set of audiovisual data. The set of text data can be produced in any appropriate manner. For

example, "production" of the set of text data may be as simple as extracting a pre-existing text transcript (e.g., a closed caption transcript) from the set of audiovisual data. Or, the set of text data can be produced from the set of audio data using a conventional speech recognition method. In step 402, the derived set of data (of the second type) is compared to the second set of data of the second type to determine the degree of similarity between the derived set of data and the second set of data. One way of making this determination is described in more detail below. In step 403, a determination is made as to whether the first set of data is relevant to the second set of data, based on the comparison of step 402. Typically, a threshold level of similarity (the expression of [the] which depends upon the method used to determine similarity) is specified so that only [a] sets of information that are sufficiently related to each other are identified as related. (This means, when the method 400 is used to generate the related secondary information region 204, that less than the allotted number of secondary information segments - or even no secondary information segments - may be displayed.)

The paragraph beginning at page 56, line 5 has been amended as follows:

The degree of similarity can be determined using any appropriate method, such as, for example, relevance feedback. In relevance feedback, a text representation of each segment to be compared (e.g., each audiovisual news story or text story) is represented as a vector, each component of the vector corresponding to a word, the value of each component being the

number of occurrences of the word in the segment. (Two words are considered identical - i.e., are amalgamated for purposes of ascribing a magnitude to each component of the vector representing the textual content of a segment - if the words have the same stem; for example, "play", "played" and "player" are all considered to be the same word for purposes of forming the segment vector.) For each pair of segments, the normalized dot product of the vectors corresponding to the segments is calculated, yielding a number between 0 and 1. The degree of similarity between two segments is represented by the magnitude of the normalized dot product, 1 representing two segments with identical words and 0 representing two segments having no matching words. The use of relevance feedback to determine the similarity between two text segments is well-known, and is described in more detail in, for example, the textbook entitled Introduction to Modern Information Retrieval, by Gerard Salton, McGraw-Hill, New York, 1983, the pertinent disclosure of which is incorporated by reference herein. Relevance feedback is also described in detail in "Improving Retrieval Performance by Relevance Feedback," Salton, G., Journal of the American Society for [information] Information Science, vol. 41, no. 4, pp. 288-297, June 1990 as well as "The Effect of Adding Relevance Information in a Relevance Feedback Environment," Buckley, C. et. al., Proceedings of 17th International Conference on Research and Development in Information Retrieval, DIGIR 94, Springer-Verlag (Germany), 1994[.], pp. 292-300, the disclosures of which are incorporated by reference herein.

The paragraph beginning at page 67, line 17 has been amended as follows:

The unsummarized text data is aligned with the unsummarized audio data. If the text data has been obtained from the audio data using a speech recognition method, then the alignment of the unsummarized text data with the unsummarized audio data typically exists as a byproduct of the speech recognition method. Otherwise, alignment is accomplished in three steps. First, the unsummarized text data is evaluated to generate a corresponding linguistic transcription network (e.g., a network describing the set of possible phonetic transcriptions). Second, a feature analysis is performed on the audio samples comprising the unsummarized audio data set to create a set of audio feature data. Third, the linguistic transcription network is compared to the set of audio feature data (using Hidden Markov Models to describe the linguistic units of the linguistic transcription network in terms of audio features) to determine the linguistic transcription (from all of the possible linguistic transcriptions allowed by the linguistic transcription network) which best fits the set of audio feature data. As a result of this comparison, the audio features of the best fit linguistic transcription are correlated with audio features in the set of audio feature data. The audio features of the best fit linguistic transcription can also be correlated with the linguistic units of the [lingusitic] <u>linguistic</u> transcription network. The linguistic units of the linguistic transcription network can, in turn, be correlated with the unsummarized text data. As a consequence of these

correlations, an alignment of the unsummarized text data with the unsummarized audio data can be obtained. Using the previously determined text summary and the alignment between the text data and audio data, an audio summary can be produced.

In the claims:

Claims 1-17 and 35-62 have been canceled.

Claim 28 has been amended as follows:

28. (Amended) A system as in Claim 18, wherein the graphical user interface includes a map region for providing a chronological description of the subject matter content of the audiovisual information and for enabling specification of control instructions that enable navigation within the audiovisual information.

Claims 63 and 64 have been added as follows:

63. (New) A system as in Claim 18, wherein the audiovisual information is represented at least partially by digital data, the means for displaying further comprising means for displaying digital data.

64. (New) A system as in Claim 18, wherein the audiovisual information is represented at least partially by analog data, the means for displaying further comprising means for displaying analog data.